MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

244 WOOD STREET

LEXINGTON, MASSACHUSETTS 02173-9108

Assistant Director
Professor of Electrical Engineering

Telephone (617) 253-3711
FAX (781) 981-6143

TO: Paul Menzel, NOAA / NESDIS

FROM: William J. Blackwell and David H. Staelin

DATE: 5 November 1997

SUBJECT: Information Content Metric for Hyperspectral Sounding Systems

INTRODUCTION

The IPO and its contractors are developing concepts for sounding systems that meet NPOESS requirements. Because these requirements are expressed in terms of environmental data records (EDR's) which link instrument design with retrieval techniques, it can be difficult to distinguish the influence of the retrieval approach from the underlying intrinsic merit of various alternative instrument configurations. This ambiguity can negatively impact engineering trade-off studies conducted by individual contractors, and can render more difficult government comparison of alternative systems approaches developed by multiple contractors. This memorandum proposes a practical information content metric which simplifies engineering concept evaluations and comparisons as a supplement to full evaluation of EDR performance.

SUMMARY OF RECOMMENDED INSTRUMENT METRIC

The proposed metric is the rms temperature and humidity profile error evaluated using linear regression retrievals based on OATs-supplied atmospheric profiles. A few thousand of these globally representative profiles should be cloud free and a comparable number should have clouds. Multi-spot retrievals would not be performed. GFE analysis software might also be provided.

DISCUSSION OF RECOMMENDED INSTRUMENT METRIC

Two different information-content metrics might reasonably be employed, one applying to cloud-free soundings and one applying to partly cloudy soundings. These are discussed separately below. These metrics do not provide an absolute measure of ultimate system performance, but rather they reasonably rank-order alternative systems with respect to the ultimate retrieval performance which might be achieved, even if that ultimate retrieval algorithm is not yet known.

One approach to estimating information content in a particular instrument data set is to determine the number of degrees of freedom in the system output which have useful signal-to-noise ratios. One technique for determining this number of degrees of freedom is to define the following linear vector representation *K* of a radiance spectrum *R*:

$$K = (QZ_N)^T L_{nn}^{-1} R$$

where *Q* is the orthonormal matrix of eigenvectors of the statistical covariance matrix associated with an ensemble of spectral radiances, L_{nn} is the noise covariance matrix, and Z_N is a square "truncation matrix" which selects *N* degrees of freedom as follows:

$$Z_{ij} = \begin{cases} \delta_{ij} \text{ if } i \le N \\ 0 \text{ Otherwise} \end{cases}$$

where *N* = number of coefficients kept. The value of *N* is chosen to ensure that all of the "information" contained in the spectral radiances is captured. The elements of the vector *K* are the eigenvalues of the "noise adjusted principal components" (NAPC) coefficients. Noise adjustment (implemented by first scaling channel gains so that all channel noise variances are equal) is particularly important for systems where the noise levels differ significantly across the channels, as they do for AIRS, ITS, and other broadband infrared spectrometers or mixed microwave-infrared systems.

These eigenvalues can be normalized such that their sum equals unity, as is done in Figure 1, where these normalized NAPC eigenvalues are plotted as a function of their sequence number (largest eigenvalue first). This figure includes only the 15-micron channels of AIRS as computed for an ensemble of 3000 NOAA / NEPC ETA-model 64-level profiles to 1 mbar for all seasons and latitudes. Note that Curve 4, corresponding to the eigenvalues of the temperature profile principal components, rapidly approaches zero near 64 coefficients, which is the number of levels employed in the radiosonde data set used.

Curve 1 presents the normalized NAPC eigenvalues for the 15-micron AIRS radiances alone. Note that the eigenvalues approach their asymptote of 2 x 10^{-6} at approximately 30 coefficients, which is then a measure of the number of useful degrees of freedom in that radiance data set. The eigenvalues are approaching an asymptote representing pure noise. If the instrument were modified so that only every fourth channel were retained, Curve 2 results, with an asymptote near 10^{-5} and 25 coefficients instead of 30. If instead the channel radiances were four times as noisy, the asymptote begins near 20 degrees of freedom and has a value near 4 x 10^{-5}. Thus we see decreasing numbers of degrees of freedom corresponds to increasing asymptotic eigenvalues and noise levels. The fact that the asymptote for the system with only every fourth channel retained lies midway between Curves 1 and 3 is consistent with the fact that reducing the ability to average data by a factor of four corresponds roughly to a doubling of the equivalent noise.

In contrast, Figure 2 presents a similar analysis for the 4-micron channels of AIRS for the same data set used before. Even though only 250 channels were used at 4 microns instead of the 475 channels employed in Figure 1 at 15 microns, the number of degrees of freedom present in the raw radiances is approximately 120, as is evident in Curve 1, for which the asymptote is approximately 1.7×10^{-10}. Curves 2 and 3 yield asymptotes near 5×10^{-10} and 2.5×10^{-9}, respectively, for systems employing every fourth channel, or all channels but with four times the noise standard deviation. The number of degrees of freedom for Curve 3 is approximately 90, and that for Curve 2 is 60; as before, more degrees of freedom correspond to higher accuracies.

It is interesting to note the effects of converting the radiances into equivalent brightness temperatures. In Figure 1, Curve 5 shows that the asymptote and number of degrees of freedom is very slightly improved using brightness temperatures at 15 microns, whereas the same operation produces a marked degradation at 4 microns, with a reduction to approximately 30 coefficients and a asymptote of 8×10^{-7}, approaching the performance of the more linear 15-micron band. In each case in Figures 1 and 2, the lower the noise level and asymptote, the greater the number of degrees of freedom.

Perhaps more important than the number of degrees of freedom, however, is the rms retrieval performance for these two systems in the cloud-free atmospheres analyzed here. The rms retrieval errors presented here are the average of the rms errors for 15 1-km thick layers. These layer errors are averages in turn of the rms errors for those of 32 levels which fall within each 1-km slab. These 32 levels are distributed up to 100-mbar pressure levels. Figure 3 presents the temperature profile retrieval error obtained using only 15-micron channels. The rms error is plotted as a function of the number of NAPC coefficients used in the subsequent linear regression which yielded the retrievals. The average rms temperature retrieval error over the atmospheric column has an asymptote of approximately 1.15 K if 22 or more coefficients are used in the regression. This asymptotic number of coefficients (22) corresponds closely to the number for the asymptote noted in Figure 1. Again, very slightly better performance is predicted if the radiances are first converted to brightness temperatures. This difference is small because the 15-micron channels are so nearly linear.

Figure 4 illustrates the same asymptotic retrieval performance for linear regressions as a function of the number of principal component coefficients used in the regression. Using the highly nonlinear 4-micron radiances, the rms asymptote has not yet been reached with 65 coefficients and 0.65 K RMS errors. Data converted to equivalent brightness temperatures reach a retrieval asymptote near 1.4 K and 19 coefficients, more nearly comparable to the 15-micron results.

Although the number of degrees of freedom is clearly related to the ultimate retrieval performance (more is better), it would seem more direct to use the linear regression retrieval performance as the metric instead. This is particularly so because the computational burden associated with evaluation of the linear retrieval performance is quite comparable to the computation required for determining the numbers of degrees of freedom from the NAPC expansion, as was done here using the data in Figures 1 and 2. Furthermore, the number of degrees of freedom alone does not indicate whether errors on some altitudes may be unacceptable. This linear retrieval approach works well for these nonlinear estimation problems because the number of channels is so large compared to the degrees of freedom in the data; the recommendations here would not apply otherwise.

Although the proposed linear regression retrieval metric permits simple and meaningful comparison of similar systems for sounding temperature profiles or other parameters in clear air, it is less useful for evaluating the relative merits of spectrometers which operate in the presence of clouds or precipitation. For example, although the 4-micron and 15-micron band responses in clear air are reasonably similar except for the accentuated temperature dependence at 4-micron wavelengths, the differential sensitivities of these two bands to clouds can be quite marked, and therefore this difference provides a means for estimating the radiance perturbations produced by such hydrometeors.

CANDIDATE INFORMATION CONTENT METRICS

In view of the foregoing, it is recommended that contractors and others be provided with an ensemble of radiances and associated atmospheric profiles which they can adapt to their own instrument spectral characteristics relative to channel definitions and noise levels. This test ensemble should include a global set of cloud-free soundings, together with a second set of cloud-perturbed soundings. The proposed metric is simply the predicted rms retrieval errors for temperature and humidity as a function of altitude for the chosen sounding ensemble, instrument channel definitions, and noise specifications. Preparation of this ensemble of soundings and predicted radiances should perhaps be a task overseen by the sounding OATs team.

DHS:emc
Attachments
pc: William L. Smith, NASA

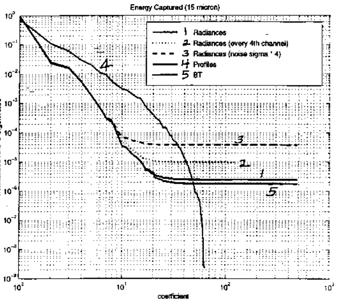

Figure 1. The NAPC transform was calculated using a simulated 475-channel 15-micron instrument (modeled after channels 1-475 of AIRS), both with and without a conversion to brightness temperatures. The energy contained in the coefficients of the GCM (ETA model) clear-air temperature profile ensemble used for the calculations is also shown. Profiles and radiative transfer models were provided by the NASA/JPL AIRS Science Team.

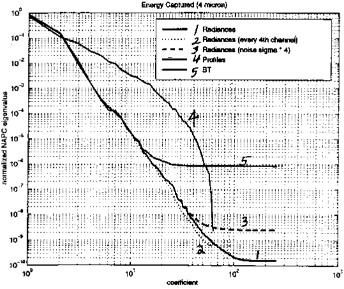

Figure 2. The NAPC transform was calculated using a simulated 250-channel 4-micron instrument (modeled after channels 1868-2118 of AIRS), both with and without a conversion to brightness temperatures. The energy contained in the coefficients of the GCM (ETA model) clear-air temperature profile ensemble used for the calculations is also shown. Profiles and radiative transfer models were provided by the NASA/JPL AIRS Science Team.

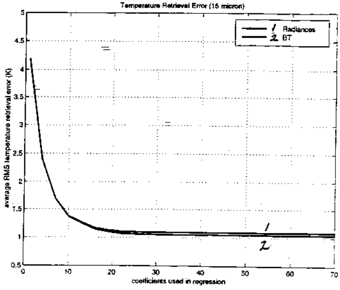

Figure 3. Linear regression was used to retrieve temperature profiles from both spectral radiances and converted brightness temperatures. Average retrieval error over the atmosphere is shown as a function of the number of NAPC coefficients used in the regression.

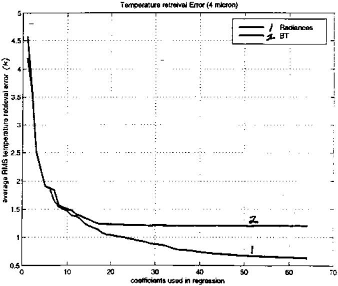

Figure 4. Linear regression was used to retrieve temperature profiles from both spectral radiances and converted brightness temperatures. Average retrieval error over the atmosphere is shown as a function of the number of NAPC coefficients used in the regression.